



4º SEMINÁRIO NACIONAL DESAPROPRIAÇÃO E REASSENTAMENTO

2025



MINISTÉRIO DOS
TRANSPORTES



IA e Raspagem de Dados na Avaliação de Imóveis

do dado bruto ao laudo confiável

Prof. Melquizedeck Souza

O que é Raspagem de Dados

Raspagem é coletar automaticamente informações públicas de sites para montar uma base organizada.

- o que aparece nos anúncios de imóveis, como preço, área, quartos e localização;
- a raspagem só lê o que já está público;
- não é invasão de sistema.

Raspagem de Dados + I.A.

A Raspagem junta as informações.

A IA ajuda a limpar, padronizar, detectar erros e explicar o que mais influencia o valor.

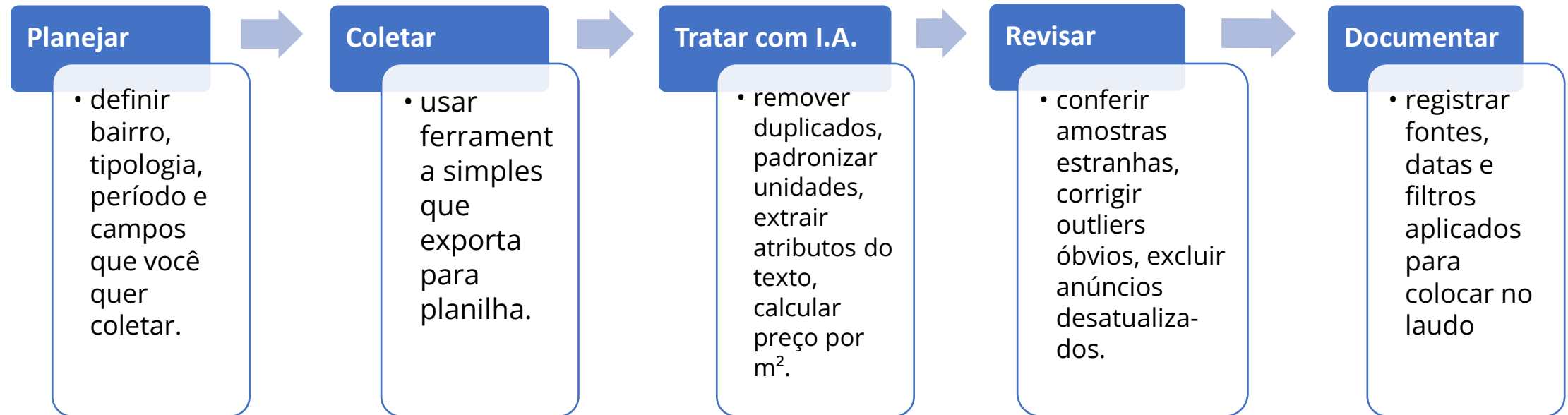
- tirar duplicados;
- completar campos ausentes;
- transformar textos em dados (exemplo: “2 vagas” vira número 2);
- apontar fatores que mais pesam no preço por m².

Quando usar Raspagem de Dados





A Raspagem de dados é utilizada quando você precisa de uma amostra sólida de comparáveis para a inferência estatística.

- coleta de anúncios do bairro ou região;
- recorte por tipologia;
- atualização da amostra quando o mercado muda;
- apoio à homogeneização por fatores ou regressão.





Como usar a Raspagem de Dados



Benefícios da Raspagem de Dados

-  **Velocidade:** horas a menos na coleta manual.
-  **Qualidade:** amostra maior e mais coerente, menos viés de seleção.
-  **Fundamentação:** gráficos e tabelas limpas que sustentam o laudo.
-  **Escala:** reaproveitamento do processo em outros bairros e tipologias.

Precauções ao usar a Ferramenta

-  **Legalidade:** seguir termos de uso do site e a LGPD; não coletar dados pessoais sensíveis.
-  **Qualidade:** remover duplicados e anúncios antigos; evitar misturar mercados diferentes.
-  **Transparência:** documentar fontes, data da coleta e filtros; anexar resumo no laudo.
-  **Ética:** citar fontes e não ultrapassar limites de uso dos portais.

Algumas ferramentas disponíveis

Sem código (clique e arrastar)

- **Octoparse:** constrói fluxos e exporta para Excel/CSV.
- **ParseHub:** bom para páginas dinâmicas; tem agendamento.
- **Web Scraper** (extensão do Chrome): mapeia páginas e exporta CSV.
- **Data Miner** / Instant Data Scraper (extensões): ótimos para páginas simples.
- **Apify Store:** “atores” prontos para vários sites; roda na nuvem.

Algumas ferramentas disponíveis

Com código (maior controle e escalabilidade)

➤ **Python**

- Requests/Httpx + BeautifulSoup4/lxml: páginas estáticas.
- Selenium ou Playwright: páginas dinâmicas (JavaScript, scroll).
- Scrapy: framework robusto para projetos grandes, com pipeline e agendamento.

➤ **Node.js**

- Puppeteer ou Playwright: automação de navegador.
- Cheerio: parser HTML rápido para páginas estáticas.

Considerações Finais

Checklist rápido:

- definir tipologia e região de pesquisa
- coletar anúncios do período mais recente (últimos 3 meses)
- remover duplicados e desatualizados
- calcular preço por m² e padronizar atributos
- registrar fontes e data de coleta para o laudo



**“Do dado bruto ao laudo confiável: ética,
método e você no comando.”**
(ChatGPT)

OBRIGADO!